SYSTEM AND METHOD FOR DYNAMICALLY EVALUATING LATENT CONCEPTS IN UNSTRUCTURED DOCUMENTS

Abstract

A system and method for dynamically evaluating latent concepts in

5   unstructured documents is disclosed. A multiplicity of concepts are extracted from a set of unstructured documents into a lexicon. The lexicon uniquely identifies each concept and a frequency of occurrence. A frequency of occurrence representation is created for the documents set. The frequency representation provides an ordered corpus of the frequencies of occurrence of each concept. A

10   subset of concepts is selected from the frequency of occurrence representation filtered against a pre-defined threshold. A group of weighted clusters of concepts selected from the concepts subset is generated. A matrix of best fit approximations is determined for each document weighted against each group of weighted clusters of concepts.